Implementing disaster tolerant and disaster recovery solutions for Microsoft SharePoint using HP storage replication technologies (Part 1)

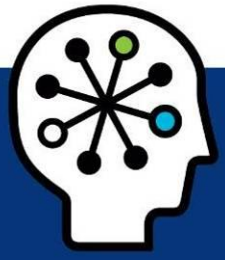


Table of contents

Introduction

Microsoft Office SharePoint Server (MOSS) is Microsoft's fastest growing technology. Companies of all sizes are leveraging its remarkable capabilities to improve employee communication and collaboration, streamline information access, and manage rapidly growing content across the enterprise.
Once companies realize the significant improvement to productivity that is gained by implementing MOSS, it soon becomes a critical business application. MOSS requires disaster tolerance and disaster recovery solutions to protect valuable information against the threat of downtime. These solutions make business operations resilient, regardless of external or internal events.

As Microsoft's largest OEM Partner, HP offers a complete portfolio of complementary technology and services to help customers get the most out of their MOSS solutions. As part this portfolio, HP offers the following products:

- HP StorageWorks Continuous Access EVA (CA EVA), which performs real-time replication between Enterprise Virtual Arrays
- HP StorageWorks Cluster Extension EVA (CLX EVA), which offers protection against application downtime from fault, failure, or site disaster by extending a local cluster between sites over metropolitan distance.

Together, these two products provide a complete disaster tolerance and disaster recovery solution that reinstates critical applications at a remote site within minutes after an adverse event.
This paper, which is the first of a two part series, focuses on implementing CA EVA and CLX EVA in a SharePoint Server 2007 scenario. Doing so provides HP disaster tolerance and disaster recovery solutions for the SharePoint application farm. This paper includes the following information:

- Provides getting started steps for implementing HP replication in existing SharePoint environments
- Showcases disaster tolerance, using a variety of disaster scenarios
- Provides how to steps for getting the SharePoint application back to a pre-disaster state, location, and resuming normal operations.

The second part of this series focuses on optimizing performance of multi-site SharePoint farms and includes a comparison between using SQL Server Replication for SharePoint data and using HP Replication Technologies.

Solution configuration

Test scenario

To demonstrate the capability of HP disaster tolerance and disaster recovery solution, testing is performed on a typical mid-range Microsoft Office SharePoint 2007 environment.

The SharePoint Server 2007 primary site application farm consists of the following three application tiers:

- Web front end servers (WFEs) (3)
- Application server (index server) (1)
- SQL Server 2008 clustered nodes, which service the application back end (2)

The primary site farm is duplicated on a secondary site. For more information, see Test Environment.

During testing, HP LoadRunner software is used to simulate1,000 SharePoint users, generating a typical collaborative workload that consists of the following:

- 40% search activity
- 20% team sites activity
- 15% document check-in check-out activity
- 10% announcement activity
- 10% event activity
- 5% My Sites activity

Approximately 0.5 TB content data is used during testing, which includes three major versions of stored content data.

The replication example has a simulated geographical distance that does not exceed 300 miles from the primary site (Site A) to the secondary site (Site B); or more specifically, a distance that has a network round-trip response time of 5 ms. An Empirix Gigabit Ethernet PacketSphere XG is used as a delay generator to simulate this distance between sites on WAN links.

When planning the solution, each tier of the SharePoint application must be considered.

WFEs

As a best practice, Microsoft recommends that there be only 1 ms delay between WFEs and the SQL Server backend. Therefore, site-to-site failovers should include the entire application farm. For example, if a failure occurs that causes the back end SQL server to fail over to Site B, the WFEs should also be redirected to operate from Site B. This redirection is typically done without human intervention when using a hardware load balancer, such as F5 BIG-IP Local Traffic Manager with WebAccelerator. For more information, see the Microsoft Technet article, Plan for front-end web server redundancy.

Index servers

The following challenges for availability exist for SharePoint 2007, which effects availability of SharePoint search functionality when site failovers occur.

- Index server redundancy cannot be achieved by installing the index role on multiple servers.
- Index server redundancy cannot be achieved by clustering index servers.

To overcome the loss of an index server, Microsoft guidance states that if the business does not require immediate search currency and availability after failover, reinstall the server and either restore from a backup or rely on slightly stale results while SharePoint search recrawls the content. For more

information about alternative methods, see "Plan and Configure availability within an Office SharePoint Server farm." During testing, recrawling content is accomplished in 40 minutes.

SQL Server back end

CLX EVA integrates with Microsoft Cluster Service and CA EVA to automate failover and failback of the SQL Server backend. This dual integration enables the cluster service to verify the status of the storage and the server cluster, which allows the correct failover and failback decisions to be made. Doing so minimizes downtime and accelerates recovery without human intervention.

The integration between these products provides the following performance benefits:

- Near instantaneous data replication
- Read/write enabling of remotely mirrored storage
- LUN presentation/mapping changes during failover
- Extensive condition checking by monitoring and recovering disk pair synchronization on an application level.

This paper documents the following test results.

- Converting an existing MOSS farm into a two-site disaster tolerant farm, as described in Test Environment.
- Validating disaster tolerance of the MOSS environment, as described in Disaster tolerance Scenarios.
- Validating disaster recovery of the MOSS environment, as described in Disaster recovery.

Test environment

Note: This is not a reference configuration, only a proof of concept.

This section describes the test environment designed to implement the solution. The servers and storage meet the requirements for the test scenario; however, based on *HP ProLiant Sizer for Microsoft Office SharePoint Server 2007* and guidance from Microsoft, the following two test-environment elements may not be appropriate for a particular production implementation:

- The test environment uses Ethernet to simulate intersite WAN links.
    - Intersite link bandwidth affects the recovery time objective (RTO) of the environment. The RTO is the length of time the business can afford to spend returning an application to operation. It includes the time required to detect the failure, to fail over the storage, and to restart the application on a new server.
    - For more information about choosing an appropriate intersite link, see the *HP StorageWorks Continuous Access EVA implementation guide*.
    - HP provides detailed guidance for designing the replication links and best practices for replication performance tuning. For more information, see the *High availability for SQL Server 2005 using array-based replication and host-based mirroring technologies* white paper.
- The test environment uses Windows 2008 network load balancing for WFEs.
    - Depending on business requirements, alternatives for load balancing WFEs include using a hardware load balancer or using virtual machines for the WFEs. Either one can be implemented as part of this solution.

- HP provides guidance for implementing virtualization in Sharepoint environments. For more information, see the *Best practices for deploying Microsoft Office SharePoint Server 2007 with Hyper-V on HP ProLiant servers* white paper.
- As part of another project, HP ran tests on Microsoft SharePoint Server 2007, using F5 hardware load balancing technology and F5 BIG-IP Local Traffic Manager with WebAccelerator, which improved throughput to a branch office by up to eight times in a symmetric configuration. For more information, see "F5 and HP Partner to Optimize SharePoint Server 2007" on the WindowsITPro website at http://windowsitpro.com/podcast/Index.cfm?fuseaction=ShowRegistration&PCID=c3df1124-8529-4c59-b35c-02016fa04621&code=EPDedf5NetworksApr09PD01082409.

Figure 1 shows a high-level view of the dual site test environment. It further shows the data path of the SharePoint application data being replicated. Four HP StorageWorks IP Distance Gateways (MPX110) provide SAN-over-WAN data transfer between HP EVA systems.

For more information about HP StorageWorks IP Distance Gateways, see the HP StorageWorks IP Distance Gateway Questions & answers website at http://h18000.www1.hp.com/products/storageworks/fcipd_gateway/qa.html?jumpid=reg_R1002_USEN.

Figure 1. High-level view of the dual site test environment

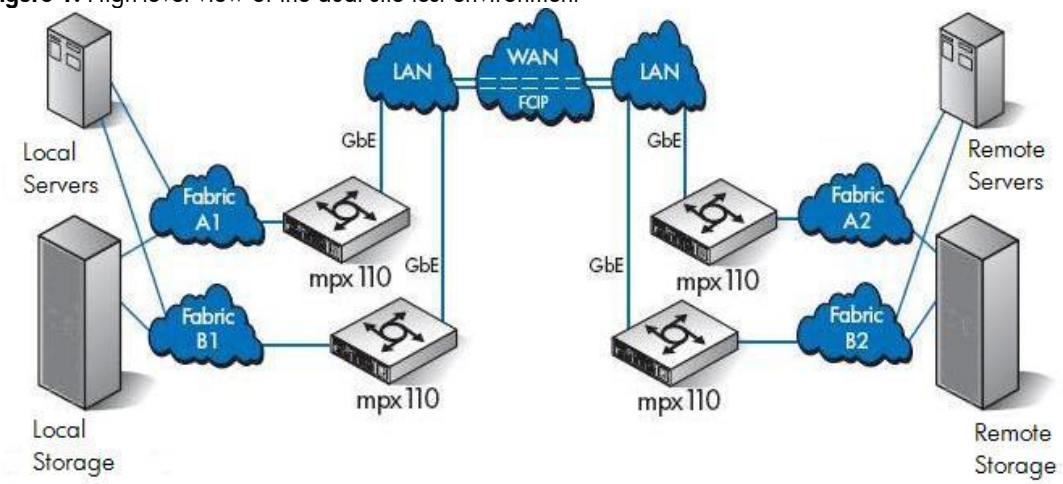


Figure 2 shows a physical view of hardware used in the dual site test environment.

Figure 2. Physical view of the dual site test environment

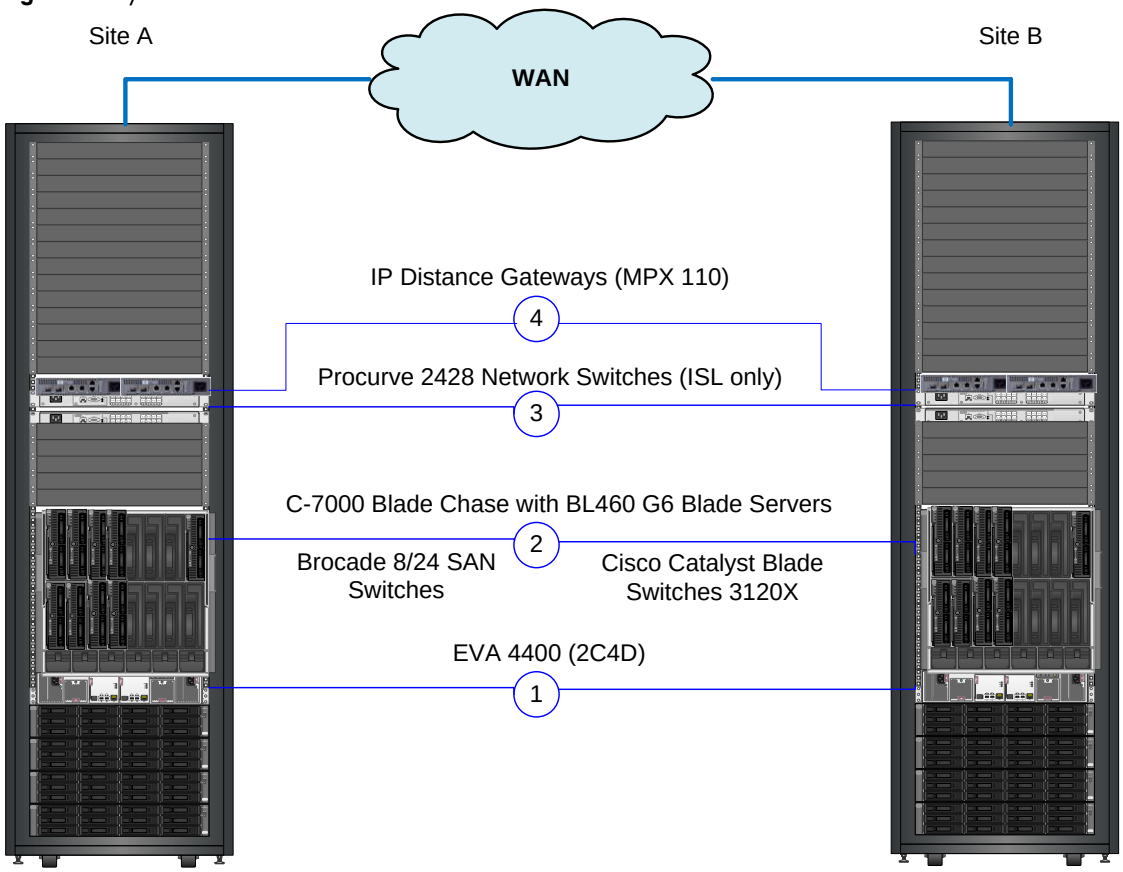


Table 1 describes the hardware, software, and firmware versions used in testing.

Table 1. Hardware and software version matrix

Hardware	Software/firmware version
SharePoint 2007 Servers: BL460 G6 Servers (8), (4 per site) WFEs per site (3) Index server per site (1) Quad-Core Intel Xeon, 2.53 GHz (2) RAM: 64 GB QLogic QMH2562 8 Gb FC HBA for HP c-Class BladeSystem (2)	 Windows Server 2008 Enterprise Edition (x64) SP2 Microsoft Office SharePoint Server 2007 SP2 HBA firmware 5.01.02 HBA Storport Driver 9.1.8.17 MPIO FF DSM for EVA 4x00/6x00/8x00 v3.03.00
SQL Servers: BL460 G6 Servers (4), (2 per site) Quad-Core Intel Xeon, 2.53 GHz (2) RAM: 64 GB QLogic QMH2562 8 Gb FC HBA for HP c-Class BladeSystem (2)	 Windows Server 2008 Enterprise Edition (x64) SP2 SQL Server 2008 Enterprise Edition (x64) SP2 HBA firmware 5.01.02 HBA Storport Driver 9.1.8.17 MPIO FF DSM for EVA 4x00/6x00/8x00 v3.03.00 Cluster Extension EVA v2.01
Storage Management Servers: BL460 G6 Servers (2), (1 per site) Quad-Core Intel Xeon, 2.53 GHz (1) RAM: 8 GB QLogic QMH2562 8Gb FC HBA for HP c-Class BladeSystem (2)	 CommandView EVA 9.0 Continuous Access EVA license enabled HBA firmware 5.01.02 HBA Storport Driver 9.1.8.17 MPIO FF DSM for EVA 4x00/6x00/8x00 v3.03.00
EVA 4400: 2C4D configuration (2) , (1 per site) 146 GB 15K FC disks per array (48)	 Firmware: XCS 09006000

Hardware	Software/firmware version
FC Switches: Brocade 8/24 SAN Switch Power Pack for HP c-Class BladeSystem switches (8), (4 per site)	Switch firmware: 6.2.0b
Network Switches: Cisco Catalyst Blade Switch 3120X for HP c-Class BladeSystem (4), (2 per site) ProCurve 2428 Network Switches (ISL only) (4)	Switch firmware: 12.2(50)SE
IP Distance Gateway: MPX 110 (4), (2 per site)	Firmware: 6.0 / software: 2.3.2.1 MPX manager: v2.0.30b127

The EVA array configuration for each site consists of 48 146-GB FC disks. Forty-six disks are used in testing, and two disks are available spares. The disk group configuration and LUN mappings are described in Table 2, Table 3, and Table 4.

Table 2. EVA disk group configuration

Disk group	Number of disks	Contents
MOSS	10	SQL install path
		Temp DB data
		Temp DB logs
Content data	12	Content DB
Logs	8	Content logs
Catalogs	16	Search catalog
		Index catalog

In accordance with Microsoft SharePoint Server best practices, SharePoint content databases are limited to 100 GB of data to enhance performance and manageability. The test environment has five content database LUNs.

Additionally, Microsoft recommends that the of SQL TempDB data files should equal the number of core CPUs, counting dual core processors as two CPUs for this purpose and each processor that supports hyper-threading as a single CPU. The test environment uses four TempDB LUNs, and four associated TempDB log LUNS.

For more information about these recommendations, see *Planning and Monitoring SQL Server Storage for SharePoint: Performance Recommendations and Best Practices*.

Table 3. SQL Server LUN mapping

Disk drive letter or mount point	Data type	Disk group	LUN size (GB)	RAID level
F:\	SQL install path SQL1	MOSS	100 GB	RAID 1
G:\	SQL install path SQL2	MOSS	100 GB	RAID 1
H:\ Temp-MP (Mount point) LUNs mounted (4) (Temp-DB1..Temp-DB4)	Data-SQL1	MOSS	1 GB Mount Point 10 GB (4)	RAID 1
I:\ Temp-DB-Logs-MP (Mount point) LUNs mounted (4) (Temp-Logs1..Temp-Logs4)	Logs-SQL1	MOSS	1 GB Mount Point 10 GB (5)	RAID 1
J:\ Temp-MP (Mount point) LUNs mounted (4) (Temp-DB1..Temp-DB4)	Data-SQL2	MOSS	1 GB Mount Point 10 GB (4)	RAID 1
K:\ Temp-DB-Logs-MP (Mount point) LUNs mounted (4) (Temp-Logs1..Temp-Logs4)	Logs-SQL2	MOSS	1 GB Mount Point 10 GB (5)	RAID 1
L:\ Content-Data (mount point) LUNs mounted (2) (CD1..CD2)	Data-SQL1	Content data	1 GB Mount Point 130 GB (5)	RAID 1 RAID 5
M:\ Content-Logs (mount point) LUNs mounted (5) (CD-Logs1..CD-Logs2)	Logs-SQL1	Logs	1 GB Mount Point 30 GB (5)	RAID 1
N:\ Content-Data (mount point) LUNs mounted (3) CD3..CD5)	Data-SQL2	Content data	1 GB Mount Point 130 GB (5)	RAID 1 RAID 5
O:\ Content-Logs (mount point) LUNs mounted (5) (CD-Logs1..CD-Logs5)	Logs-SQL2	Logs	1 GB Mount Point 30 GB (5)	RAID 1
P:\ MSDTC	Logs	MOSS	1 GB	RAID 1

Table 4. SharePoint Server LUN mapping

Disk drive letter or mount point	Data type	Disk group	LUN size (GB)	RAID level
E:\ Presented to Index Server	Search index	Catalogs	200 GB	RAID 1
E:\ Presented to WFE-1	Copy of search catalog data	Catalogs	200 GB	RAID 1
E:\ Presented to WFE-2	Copy of search catalog data	Catalogs	200 GB	RAID 1
E:\ Presented to WFE-3	Copy of search catalog data	Catalogs	200 GB	RAID 1

Implementing data replication with Continuous Access EVA

This section describes the implementation tasks that are performed to enable data replication using Continuous Access EVA.

The following is a checklist of activities for implementing CA EVA:

- Verify EVA setup
  - Ensure all cabling requirements are met when connecting each array to the fabric
  - Install replication licenses
- Set Fibre Channel switch configuration settings for replication
- Create fabrics and zones
- Create data replication groups (DR groups)
  - Create disk array groups on the remote site EVA, mimicking the local site EVA, so that they are selectable during the creation process
  - Select the desired replication mode
  - Select the DR group write history log location and size
  - Present the destination virtual disks

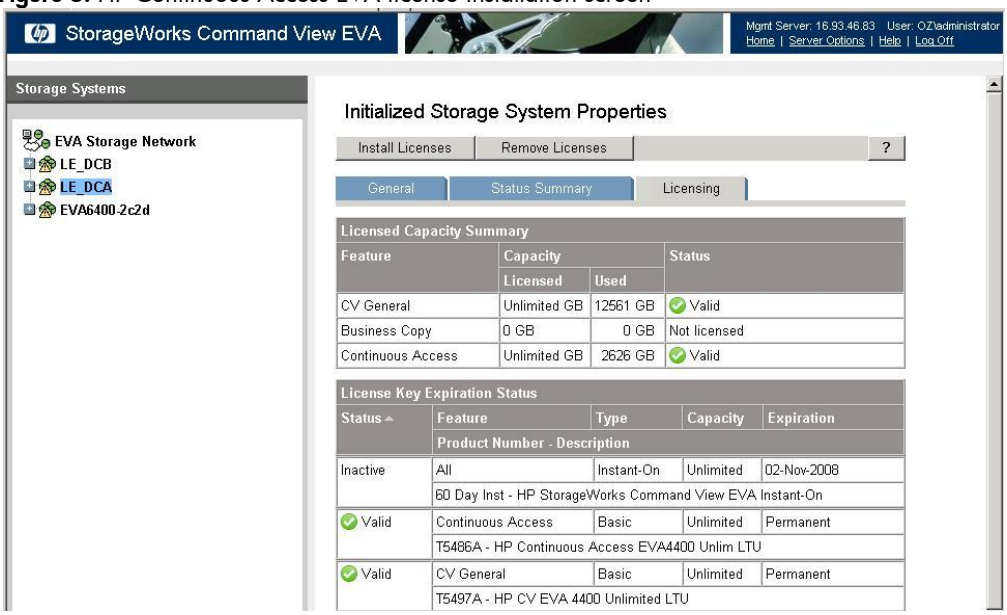
Verify EVA setup

Verify the cabling between the arrays and Fibre Channel switches meets remote replication requirements. The supported cabling scheme depends on the array controller hardware and software features. For a description of all cabling options, along with best practices for cabling, see the *HP StorageWorks SAN design reference guide*.

HP Continuous Access EVA software is installed as part of the Command View EVA software suite installation but is disabled by default. Enable CA EVA functionality by simply installing a Continuous Access EVA replication license for each array (local and remote) in a remote replication relationship. Replication licenses are based on the amount (in TB) of replicated data on each array.

Figure 3 shows where HP Continuous Access EVA license are installed using Command View EVA.

Figure 3. HP Continuous Access EVA license installation screen



Set the fibre channel switch configuration settings for replication

All FC switches that are in the path from the source array to the destination array must be configured to work in a CA EVA environment. For B-Series switches, access each switch by using Telnet, and configure the switch as follows:

- switchdisable
- iodset (enables In Order Delivery)
- aptpolicy 1 (enables port-based routing)
- portCfgISLMode [slot/port] 1 (set for all MPX110 switch ports)
- portCfgGport [slot/port] 1 (set for all MPX110 switch ports)
- switchenable

Creating fabrics and zones

When implementing CA EVA with CLX EVA, follow the zoning guidelines described in the Implementing Disaster Tolerance with Cluster Extension EVA section of this document.

Create data replication group(s)

A data replication group (DR group) is a CA EVA software connection between source virtual disks and destination virtual disks. A DR group can contain virtual disks from multiple array disk groups, but all DR group member virtual disks must be in the same array and must be set to use the same preferred controller on the array.

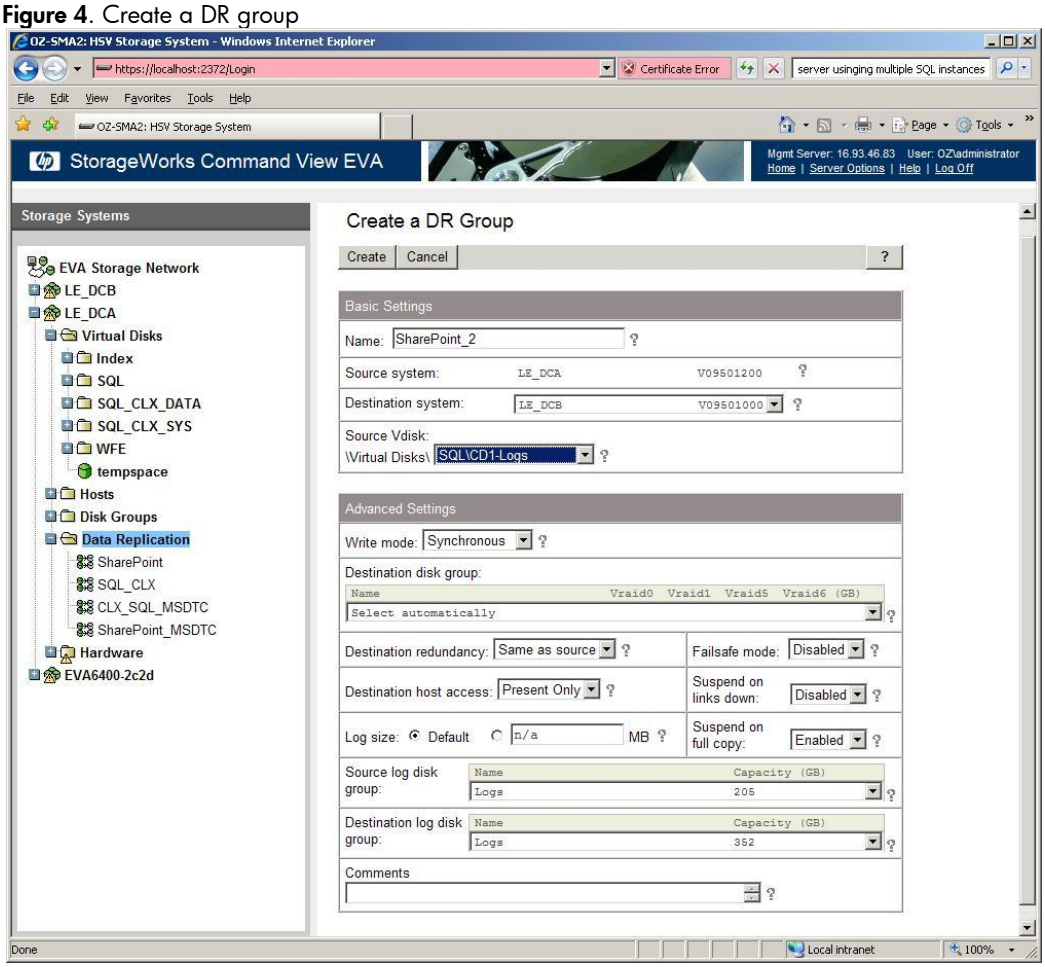
HP Continuous Access EVA requires that all members (virtual disks) of a DR group be managed by the same storage array controller. By default, one host port on the managing controller is used for replication traffic.

HP recommends that all virtual disks used by a single application be in a single DR group. Only one application per DR group is recommended. However, remember that applications, such as SQL Server, can be implemented by using multiple application server instances.

In our test environment, a second instance of SQL Server 2008 is used to store half of the SharePoint content databases. By dividing the SharePoint farm into independent web applications, a second DR group can be created that allows even balancing of the application workload across both the storage controllers. Using two DR groups provides a second bidirectional replication path, which improves the SharePoint application and replication performance.

Use Command View EVA management software to create the DR groups with only a few clicks. Figure 4 shows a DR group being created using Command View EVA.

Figure 4. Create a DR group



As part of the DR group creation process, the first virtual disk is added to the DR group. Therefore, it is best to create the disk groups on the destination array, mimicking the source environment, before creating the DR group. Then, as members are added to the DR group, the appropriate destination disk group can be selected.

During creation of the DR group, most options are self-explanatory, such as selecting the source and destination arrays. Options for the write mode, failsafe mode, and replication log are discussed below.

Write mode option

The DR Groups write mode is specified during the creation of a DR group. The options are synchronous and asynchronous. Testing is performed by using synchronous replication write mode, which means that the array acknowledges I/O completion after the data is cached on both the local and destination arrays.

Failsafe mode option

Failsafe mode is only available when a DR group is being replicated in synchronous mode. It specifies how host I/Os are handled if data cannot be replicated between the source and destination array. Testing is performed with Failsafe mode disabled.

The failsafe options are as follows:

- Failsafe enabled: All host I/O to the DR group is stopped if data cannot be replicated between the source array and destination array. This ensures that both arrays always contain the same data (RPO of zero). A failsafe-enabled DR group can be in one of two states.
    - Locked (failsafe-locked): Host I/O and remote replication stops because data cannot be replicated between the source and destination array.
    - Unlocked (failsafe-unlocked): Host I/O and remote replication resumes once replication between the arrays is re-established.
- Failsafe disabled: If replication of data between the source and destination array is interrupted, the host continues writes to the source array, but all remote replication to the destination array stops and I/Os are put into the DR group write history log until remote replication is re-established.

Replication log options

The size of the DR group write history log and the disk group locations on source and destination arrays are specified during DR group creation. By default, the write history log is 100 GB for each DR group. There is a write history log location for both source and destination arrays, because the source and destination roles are reversed when a failover occurs.

The DR group-write history log stores data when replication to the destination DR group is stopped- This occurs because the destination DR group is unavailable or suspended. The process is called logging. When replication resumes, the contents of the log are sent to the destination virtual disks in the DR group. This process of sending I/Os contained in the write history log to the destination array is called merging. Because the data is written to the destination in the order that it is written to the log, merging maintains an I/O-consistent copy of the DR group data at the destination.

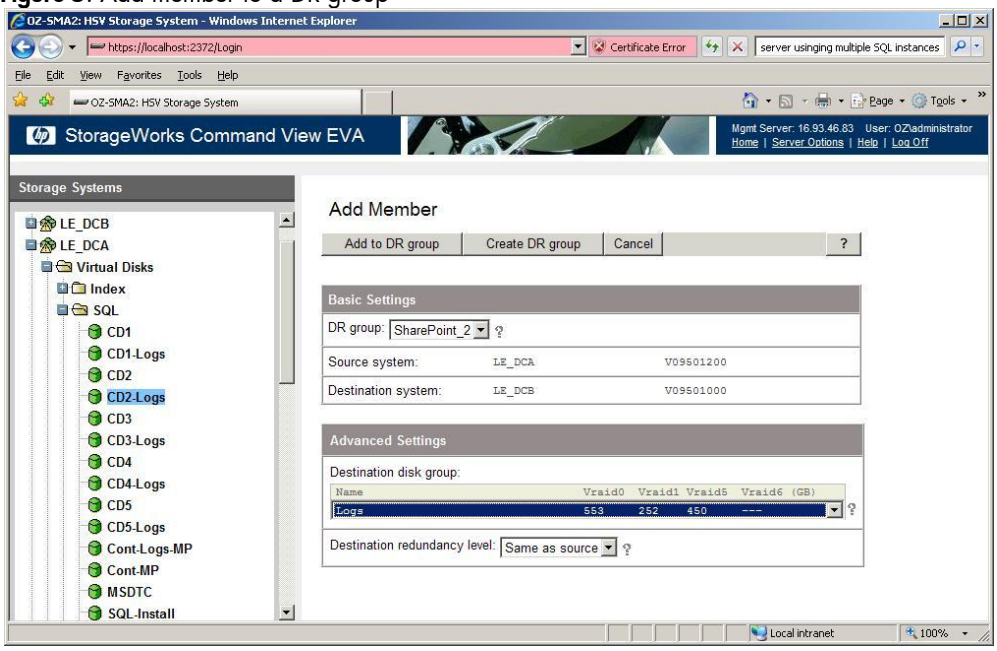
Normalization

The method of synchronizing source and destination virtual disks is called normalization. When a DR group is first created, or when a new member is added to the DR group, a full-copy normalization occurs. All data on each Vdisk member of the DR group is then copied from the source array to the destination array, bringing the two arrays into synchronization. Normalizations copy data from the source array to the destination array in 128 KB blocks. Even though the array prioritizes front end I/O requests over normalization activity, a performance impact can be incurred when adding multiple Vdisks to a DR group; therefore it is best to create the DR groups during off-peak hours of operation.

Normalization also occurs when a write history log overflows. This may happen if the replication link is down for a longer time than the log size can accommodate. When a write history log overflows, the controller invalidates the log contents and marks the DR group for normalization. In some cases, normalization will be optimized to copy only blocks that are written before the write history log overflows, not a full-copy normalization. The array tries to optimize the full copy whenever possible.

Once the DR group is created, additional virtual disks can be added to the DR group by selecting the virtual disk and by using the Data Replication tab on the Vdisk properties, as shown in Figure 5.

Figure 5. Add member to a DR group



Present the destination virtual disks

Once all the Vdisks are added to the DR group, the Vdisk replicas on the destination array must be presented to the application servers on the remote site. The replicas are marked as read-only to the remote hosts until failover occurs.

After the Vdisks are presented to the remote hosts, use Windows Failover Cluster Manager add node functionality to join the remote nodes to the SQL Server cluster.

For more information, see the *HP StorageWorks Continuous Access EVA implementation guide*.

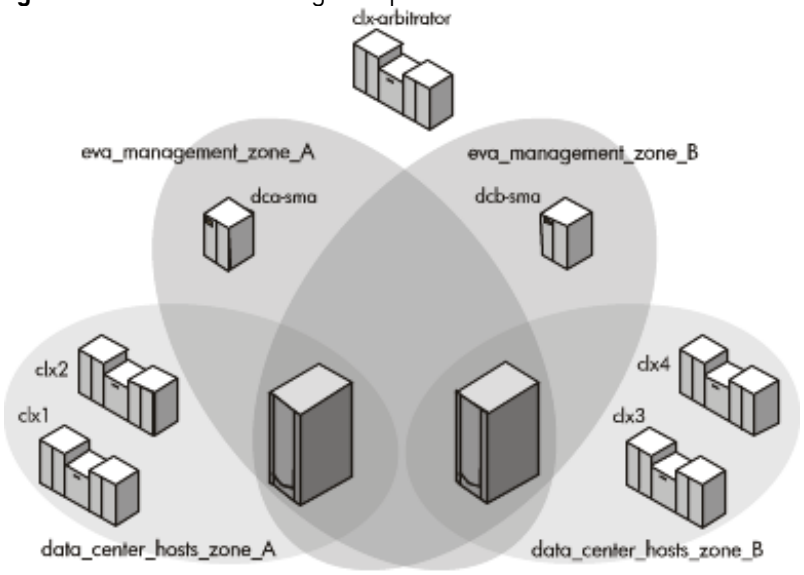
Implementing disaster tolerance with Cluster Extension EVA

Cluster Extension EVA integrates with the Microsoft Cluster Service and Continuous Access to allow access to the remote data copy if an application failover occurs. In a SharePoint environment, it is installed on each of the SQL Server cluster nodes and is configured by adding CLX as a cluster resource to each SQL Server application cluster group from within the Windows Failover Cluster Manager. CLX installation and configuration can be performed in just a few minutes and does not require application downtime for installation; however, during testing, the application is unavailable for approximately 10 minutes while the application cluster groups are failed over to each of the four SQL Server cluster nodes to verify installation.

Prerequisites

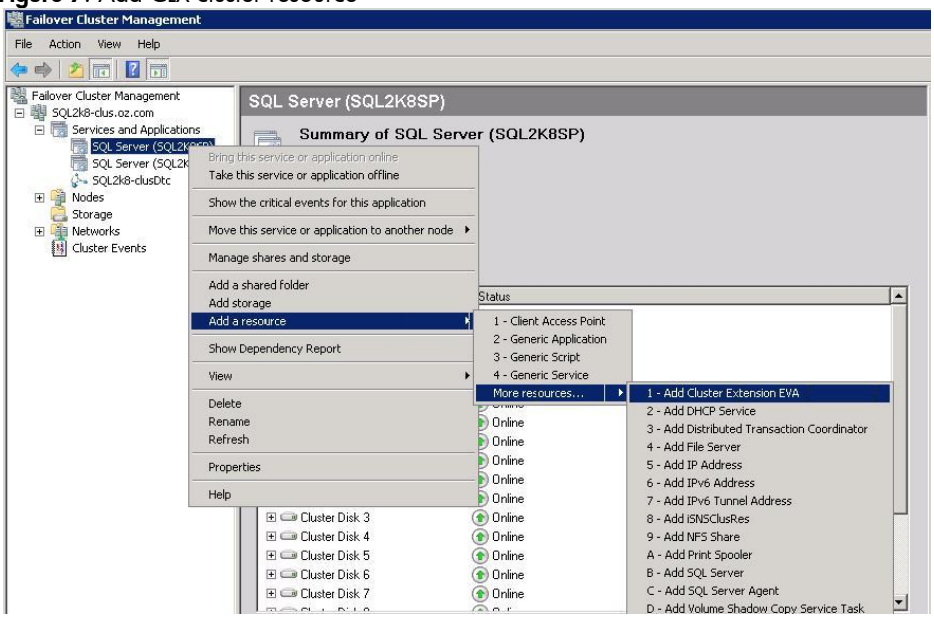
- Ensure that the following conditions are met before installing EVA Cluster Extension Software:
  - Continuous Access EVA links are redundant and bidirectional.
  - Alternate I/O paths between servers and the EVA storage system are set up.
  - The cluster nodes in your zoning configuration each recognize their respective (local) EVA, but not the remote EVA. Each management server recognizes both EVAs and can recognize each other. Figure 6 shows a zoning example.

Figure 6. Fibre Channel zoning example



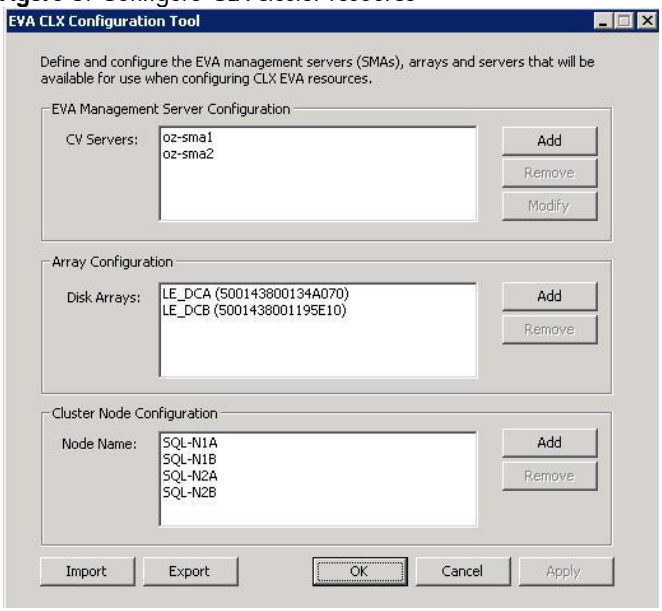
• After ensuring that the prerequisites are met and installing the CLX software on each SQL Server cluster nodes, add a CLX cluster resource to each SQL Server application cluster group as shown in Figure 7.

Figure 7. Add CLX cluster resource



• After adding the CLX cluster resource, select the resource properties and configure it as shown in Figure 8.

Figure 8. Configure CLX cluster resource



• At this point SQL Server can be installed on the remote SQL server nodes, using the add node to existing SQL Server installation option.

• Next, complete the CLX configuration by adding the CLX cluster resource as a dependency of each disk resource in the application cluster group in Windows Failover Cluster Manager.

Note: If the CLX cluster resource is added as a dependency to the cluster disks before SQL Server is installed on the remote site cluster nodes, SQL will recognize the disk dependency and fail the cluster resource check during installation.

• Finally, install SharePoint Server 2007 and the appropriate Service Packs on the remote WFEs and Application Servers, allowing the remote servers to join the SharePoint farm.

For more information about CLX, see the *EVA Cluster Extension Software installation guide* and the *EVA Cluster Extension Software administrator guide*.

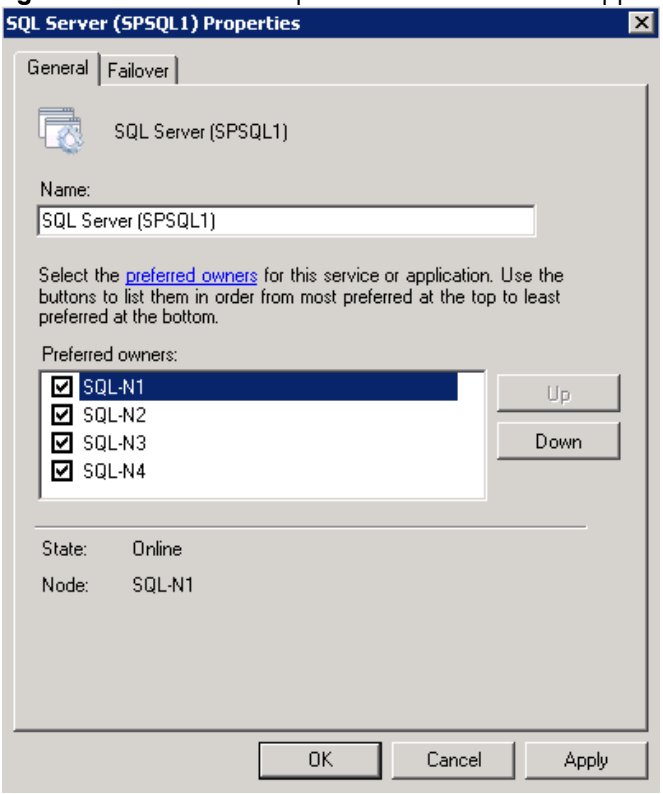
Disaster tolerance scenarios

This section describes the disaster tolerance scenarios used in testing, including definitions of configurable options that control the behavior of failover events. Using the examples in this section, administrators may configure available options specifically for their business requirements.

Site A cluster node failure

The test environment uses Windows Failover Cluster Manager to set the order of preferred owners for SharePoint's SQL back end, as shown in Figure 9. In the event of a Site A cluster node failure, the applications will fail over to the remaining local (Site A) SQL Server node before attempting a site-to-site failover. A local failover occurs within seconds.

Figure 9. Set the order of preferred owners in SQL application cluster



Site A all node failure

When all cluster nodes in Site A fail, the node failures trigger the SQL Server cluster to fail over to Site B. During testing, this failure is created ejecting both Site A SQL Server nodes.

During the failover process, CLX EVA cluster resource for the SQL Server disk resources fails over the Continuous Access EVA data replication direction for the DR group. The result is that the DR group on Site B is placed in source mode and the DR group on Site A is placed in destination mode. Once SQL Server is back online, transactions can resume, with all data now replicating from Site B to Site A. For this type of failover scenario, the user-defined parameters for the HP CLX EVA resource have no effect on failover behavior or takeover action.

In the test environment, WFEs on Site B are configured as SharePoint search query servers. A copy of the search index is replicated to each WFE, including WFEs on Site B, and they are able to provide some search capability for the site. However, until the Index server is restored, and the content is recrawled, the search data does not include new content. Note that the WFEs on Site B are paused in NLB until site failover. In this configuration, WFEs on Site B do not service the SharePoint front end when SharePoint is operating from Site A. This can be similarly implemented using a hardware load-balancing solution or by using virtual WFE servers.

ISL link failure

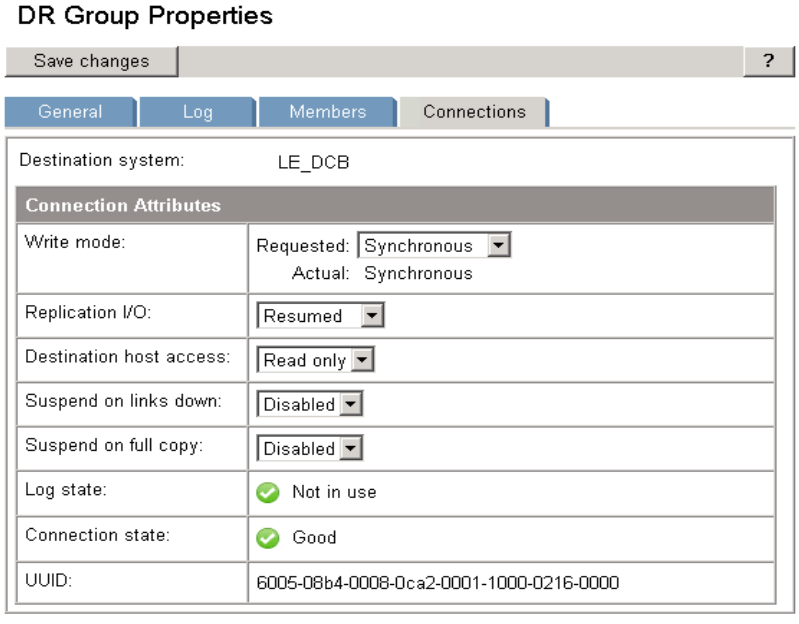
When an ISL failure occurs, the behavior depends on the DR group *Suspend on links down* property.

For application high availability, in most cases it is best to set the *Suspend on links down* property to disabled, which allows the SharePoint application to stay online in Site A in the event of an ISL failure. If the *Suspend on links down option* is enabled the SQL back end, and SharePoint, is suspended until replication can be resumed.

If *Suspend on links down* is disabled, the data replication group goes into logging state, because there is no replication communication between the two arrays. All subsequent I/Os are copied to the data replication group write history log until the log becomes full, or until the replication link between the arrays is restored before the log becomes full. If the write history log becomes full, the data replication group is marked to do a full copy from Site A to Site B when connectivity is restored. If the link is restored before the data replication group is marked for full copy and resynchronizes, a delta (merge) resynchronization occurs rather than a full copy.

The *Suspend on links down* property is accessed selecting the DR group properties using Command View EVA, as shown in Figure 10.

Figure 10. Suspend on links down DR group property



Site A access to storage failure

When shared storage resources being used by SQL Server become unavailable, a cluster failover event is triggered. Once the cluster node running SQL Server realizes the disk resources are unavailable, Microsoft failover cluster attempts to move the SQL Server instance to the second node in the list of preferred cluster nodes. If the second node in the list is the same site as the node previously running site, the HP CLX EVA resource detect that the local array is no longer available. In most cases, the desired behavior is to automatically fail over to Site B.

HP CLX EVA can only consistently, automatically fail over if the parameter *Use non-current data OK* is set to **YES** in the Cluster Extension EVA resource property.

The *Parameter Use non-current data OK* specifies whether or not the EVA Cluster Extension can fail over to the destination site in cases where the data at the destination site might not be current. This might occur in the unlikely event that an HP Continuous Access EVA link goes down prior to the fail over to the remote site:

- Setting this value to **YES** means that regardless of whether or not the data is current in the remote site, the resource comes online.
- Setting the value to **NO** means that if the data is current in the remote site, the resource comes online even though *Use non-current data OK* is set to **NO**. If the data is not current in the remote site, or if the array cannot determine if data is current, the resource will not come online.

Once SQL Server successfully fails over and is online in Site B, transactions can resume. At this point, the data replication group goes into logging state, because there is no replication communication between the two arrays. All subsequent I/Os are copied to the data replication group write history log until the log becomes full, or until the replication link between the arrays is restored with resynchronization before the log becomes full. If the write history log becomes full, the data replication group is marked to do a full copy from Site B to Site A. If connectivity between the two arrays is restored before the data replication group is marked for full copy, a delta (merge) resynchronization occurs rather than a full copy.

Site A site failure

In the event of a full-site failure, the behavior is the same as in the Site A access to storage failure, with the exception that the cluster will immediately try to bring the SQL Server instance online in Site B. Automatic recovery of the clustered SQL Server instance once again relies on the Cluster Extension EVA parameter *Use non-current data OK* for Microsoft Failover Cluster to successfully fail over the SQL Server instance to Site B.

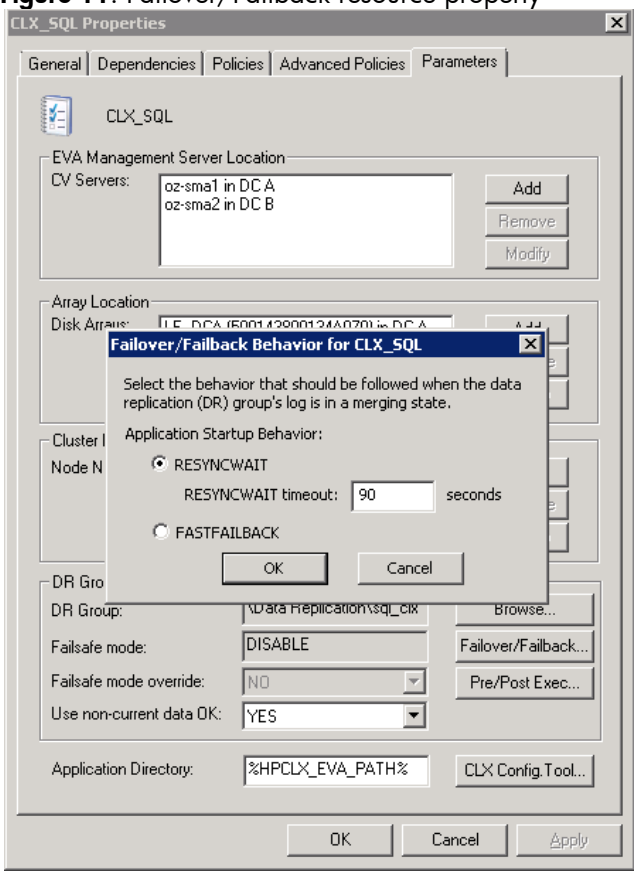
Failover during DR group resynchronization

CLX EVA has a resource property that determines Failover/Failback behavior that occurs when a site failover occurs during resynchronization.

If the Failover/Failback property of the HP CLX EVA resource is set to *RESYNCWAIT,* HP EVA waits until the resynchronization completes before successfully coming online. To ensure automatic failover, *RESYNCWAIT* should be set, even though it is not the default value. *FASTFAILBACK* is only useful if there is connectivity between the two arrays, which may not be the case for all disaster scenarios.

Figure 11 shows the location of the HP Cluster Extension resource property page under the Parameters tab.

Figure 11. Failover/Failback resource property



Disaster recovery

Cluster node recovery (failback)

Because there is no disruption to the SAN or intersite links in either the Site A single node failure scenario or the Site A all node failure scenario, the SQL Server application can be manually moved back to the original nodes on Site A. Once the nodes in Site A are back online, you can manually move the SQL Server application by using Windows Failover Cluster Manager and HP CLX EVA.

Manual failback is performed by using Windows Failover Cluster Manager, which takes the SQL Server instance offline and brings it back online on Site A. During testing, it took less than 1 minute to perform manual failback.

For maximum SQL Server instance availability, allow resynchronization to complete before attempting to move the SQL cluster group back to Site A. Otherwise, the SQL Server instance remains offline until resynchronization completes.

During the failover operation, HP CLX EVA reverses the data replication direction of the disk resources to replicate from Site A to Site B, thus returning the source/destination relationship of the DR group to its pre-failure state.

ISL link connectivity resumed

When an ISL link is restored, data replication automatically resumes if the suspend on links down parameter is disabled. If the write history log becomes full, the data replication group is marked to do a full copy from Site A to Site B, which occurs when connectivity is restored. If the link is restored before the data replication group is marked for full copy and resynchronizes, a delta (merge) resynchronization occurs rather than a full copy.

If the suspend on links down option is enabled, it is necessary to manually resume replication by setting the Replication I/O DR property to resumed, as shown in Figure 10.

Site A access to storage resumed

When connectivity to the storage on Site A is restored, the EVA in Site A will show the DR group mode in the destination state after a site failover to Site B has occurred. However, depending on the nature of the storage failure, in some cases the data replication group on both sites may be in *source* state in some cases.

When this happens, the steps to perform the failback operation depend on the settings of the DR group within HP Command View EVA.

If Suspend on links down is set to Enabled, the data replication group needs to be manually resynchronized by resuming replication I/O before it can fail back. This is achieved by setting the *Replication I/O* parameter to *Resumed*, as shown in Figure 10.

Once replication resumes, the data replication group changes from *source/source* mode to *source/destination* mode, with source in Site B replicating to the destination volumes in Site A. A full-copy or merge resynchronization takes place, as previously described.

For maximum SQL Server instance availability, wait until resynchronization completes before attempting to move the SQL cluster group back to Site A, using Windows Failover Manager. Otherwise, SQL Server and SharePoint will remain offline until resynchronization completes.

Site A failback

Failback to Site A is performed identically to the steps described in *Site A access to storage resumed*, with the exception that the cluster nodes in Site A must also be brought online before attempting a failback operation.

Additional steps must be taken to complete returning SharePoint to its original state and location. As previously stated, the following functions must be performed:

- WFE redirection
- Restore the index server role to the Site A index server
- Recrawl the content databases

During testing, these steps are performed in 56 minutes, using online backups.

Failback during resynchronization

The recovery process for failback during resynchronization is identical to the Site A cluster node recovery. SQL Server can be manually moved back to Site A by using Microsoft Failover Cluster Manager. If the Failover/Failback property of the HP CLX EVA resource is set to *RESYNCWAIT,* HP EVA waits until the resynchronization completes before successfully coming online on Site A. Otherwise, it is best to wait until resynchronization is complete before attempting a fail back to Site A.

Summary of key findings

The following is summary of key findings discovered during testing.

- Appropriate sizing, including hardware and replication links, is critical for success
- Plan for using two SQL instances for the SharePoint environment, each assigned to their own DR group, to improve application and replication performance
- Validate cabling and switch zoning configuration
- Set Fibre Channel switch configuration settings for replication, to guarantee in-order packet delivery
- Use proper configuration and load order when implementing HP storage replication products, to ensure a smooth and successful implementation
    - Enable HP Continuous Access EVA licenses
    - Create array disk groups on the remote EVA
    - Create DR groups
    - Add Vdisks to DR groups
    - Present replica disks to remote hosts
    - Install HP Cluster Extension EVA
    - Create a HP Cluster Extension EVA cluster resource for each SQL cluster instance
    - Add remote nodes to the cluster
    - Install SQL on remote cluster nodes, using the using the add node to existing SQL Server installation option
    - Add HP Cluster Extension EVA cluster resource as a dependency for each disk resource in the SQL Server instance's cluster group, completing the HP CLX configuration.
- Create and populate DR group during low activity times to avoid a performance impact during replication of the initial full copy of LUNs
- Configure the following parameters that determine failover/failback behavior, according to the appropriate business requirements:
    - Use non-current data OK
    - Suspend on links down
    - *RESYNCWAIT/FASTFAIBACK*

Conclusion

As part of the HP SharePoint portfolio, HP offers StorageWorks Continuous Access EVA data replication technology and HP StorageWorks Cluster Extension EVA that provide real-time protection against application downtime from fault, failure, or site disaster. These two products are easily implemented. Together, they provide a complete disaster tolerance and disaster recovery solution to reinstate critical applications at a remote site within minutes after an adverse event.

For more information

- Disaster-proof solutions from HP
  http://h71028.www7.hp.com/enterprise/cache/523434-0-0-0-121.html
- Microsoft Storage Solutions from HP
  www.hp.com/storage/microsoft
- HP Continuous Access EVA documentation and information
  http://h18006.www1.hp.com/products/storage/software/conaccesseva
- HP Cluster Extension EVA documentation and information
  http://h18000.www1.hp.com/products/storage/software/ceeva/index.html
- HP Customer Focused Testing solutions
  http://h71028.www7.hp.com/enterprise/cache/599679-0-0-225-121.html

Call to action

- www.hp.com/go/SharePoint
- www.hp.com/go/Storage
- www.hp.com/go/EVA

Technology for better business outcomes

4AA2-8574ENW, September 2009